

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 1, 2001, 15:52:42 ; Search time 170.72 Seconds
(without alignments)
18.693 Million cell updates/sec

Title: US-09-331-631A-8_COPY_33_79

Perfect score: 275
Sequence: 1 GDDDPKRYEDCRRCCEWDT.....OCESCKSQYGEKDDQQRHR 47

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 195891 segs, 67900655 residues

Total number of hits satisfying chosen parameters: 195891

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: pir66:*
2: pir1:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	275	100.0	588	1 FMCNAB	alpha-globulin B p
2	140	50.9	505	2 S06398	vicilin precursor
3	124	45.1	566	2 S22477	alpha-globulin typ
4	94.5	34.4	509	2 S08059	protein PVI00 (imp
5	90	32.7	810	2 T44430	arginine/glutamate
6	75	27.3	47	2 JC5557	variant surface gl
7	67.5	24.5	471	2 A35480	hypothetical prote
8	65.5	23.8	411	2 T29475	Machado-Joseph dis
9	62	22.5	360	2 S50830	co-chaperone and h
10	61	22.2	189	2 F64533	hypothetical prote
11	61	22.2	265	2 A33513	multidrug resistanc
12	61	22.2	1905	2 T18267	hypothetical prote
13	60.5	22.0	897	2 T21688	cag island protein
14	60.5	22.0	1819	2 A71928	hypothetical prote
15	60	21.8	241	2 H81536	hypothetical prote
16	60	21.8	456	2 B72130	frame-shift with c
17	60	21.8	774	2 H81540	hypothetical prote
18	60	21.8	811	2 E72003	hypothetical prote
19	60	21.8	1927	2 G64585	cag pathogenicity
20	59.5	21.6	122	2 C53234	globulin-10 - maiz
21	59.5	21.6	230	2 C72256	inf-related prote
22	59.5	21.6	407	2 T02258	globulin1 - maize
23	59.5	21.6	540	2 S21825	vicilin-like stora
24	59.5	21.6	573	2 A53334	globulin-1S, GDB1S
25	59.5	21.6	582	2 B53334	vicilin-like stora
26	59.5	21.6	929	2 S75098	hypothetical prote
27	59.5	21.6	1234	2 T30160	hypothetical prote
28	59	21.5	85	2 E42825	Kruppel-type zinc
29	59	21.5	491	2 S52920	disintegrin (EC 3.

30	59	21.5	544	2 S52477	disintegrin (EC 3.
31	59	21.5	748	2 S66129	disintegrin (EC 3.
32	59	21.5	1339	1 S20052	DNA-directed DNA p
33	59	21.5	2150	2 S71629	sensory transducti
34	59	21.5	2165	2 T21371	hypothetical prote
35	58.5	21.3	1085	2 S62516	hypothetical coile
36	58	21.1	338	2 S04321	legumin B (clone p
37	58	21.1	419	2 T18450	hypothetical prote
38	57.5	20.9	225	2 B35387	hypothetical prote
39	57.5	20.9	527	2 F64210	hypothetical prote
40	57	20.7	330	2 B71625	hypothetical prote
41	57	20.7	342	2 JC7110	ritin PEB0040c - m
42	57	20.7	493	2 T01495	brain-specific mem
43	57	20.7	942	2 T02446	hypothetical prote
44	57	20.7	1898	1 A45973	trichohyalin - hum
45	56.5	20.5	388	2 T31887	hypothetical prote

ALIGNMENTS

RESULT 1

FMCNAB
alpha-globulin B precursor (clone C72) - upland cotton

N:Alternate names: seed storage protein; vicilin precursor
C:Species: Gossypium hirsutum (upland cotton)

C>Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 16-Jul-1999
C:Accession: A30838; S06911

R:Chlan, C.A.; Pyle, J.B.; Legocki, A.B.; Dure III, L.
Plant Mol. Biol. 7, 475-489, 1986

A:Title: Developmental biochemistry of cottonseed embryogenesis and germination XVIII
A:Reference number: A30838

A:Accession: A30838
A:Molecule type: mRNA

A:Residues: 1-588 <CH2>

A:Cross-references: GB:ML6891, NID:q167374; PIDN:AAA33071.1; PID:q167375
A:Experimental source: var. Coker 201

R:Chlan, C.A.; Borroto, K.; Kamalay, J.A.; Dure III, L.
Plant Mol. Biol. 9, 533-546, 1987

A:Title: Developmental biochemistry of cottonseed embryogenesis and germination. XIX.
A:Reference number: S06398

A:Accession: S06398
A:Status: not compared with conceptual translation
A:Molecule type: DNA

A:Residues: 1-81 <CH2>
A:Comment: This is a seed storage protein.

C:Superfamily: glycinin
C:Keywords: glycoprotein; seed; storage protein

F:1-25/Domain: signal sequence #status predicted <Sig>
F:26-588/Product: alpha-globulin storage protein #status predicted <MAT>

F:417/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 100.0%; Score 275; DB 1; Length 588;
Best Local Similarity 100.0%; Pred. No. 1.2e-22;

Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GDDDPKRYEDCRRCCEWDTGKQEOQCCEESCKSQYGEKDDQQRHR 47
Db 33 GDDDPKRYEDCRRCCEWDTGKQEOQCCEESCKSQYGEKDDQQRHR 79

RESULT 2
S06398
alpha-globulin type A precursor - upland cotton

N:Alternate names: seed storage protein
C:Species: Gossypium hirsutum (upland cotton)

C>Date: 31-Mar-1990 #sequence_revision 31-Mar-1990 #text_change 30-Sep-1993
C:Accession: S06398

R:Chlan, C.A.; Borroto, K.; Kamalay, J.A.; Dure III, L.
Plant Mol. Biol. 9, 533-546, 1987

A:Title: Developmental biochemistry of cottonseed embryogenesis and germination. XIX.
A:Reference number: S06398

A:Residues: 1-471 <RED>
 A:Cross-references: GB:M33823; NID:g162487; PID:g162488
 C:Superfamily: variant surface glycoprotein
 C:Keywords: glycoprotein

Query Match 24.5%; Score 67.5; DB 2; Length 471;
 Best Local Similarity 31.0%; Pred. No. 3.4;
 Matches 13; Conservative 11; Mismatches 17; Indels 1; Gaps 1;

QY 3 DDPRRRCCEMDTRGQKEQOQCEESCKSQYGEKDDQ 44
 DB 402 NDKAKETE-CNSPCWKDEKDEKCKISECKQAEKNEQ 442

RESULT 8
 T29475
 hypothetical protein T01D1.6 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jan-2000
 C:Accession: T29475
 R:Bradshaw, H.; Wohldmann, P.
 A:Description: The sequence of C. elegans cosmid T01D1.
 A:Reference number: Z20623

A:Accession: T29475
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-411

 A:Cross-references: EMBL:U80455; PIDN:AAB37887.1; GSPDB:GN00020; CESP:T01D1.6
 A:Experimental source: strain Bristol NZ; clone T01D1
 C:Genetics:

A:Gene: CESP:T01D1.6
 A:Map position: 2
 A:Introns: 25/3: 304/3
 C:Superfamily: gliadin

Query Match 23.8%; Score 65.5; DB 2; Length 411;
 Best Local Similarity 35.9%; Pred. No. 5;
 Matches 14; Conservative 11; Mismatches 11; Indels 3; Gaps 2;

QY 10 EDCRRRCCEMDTRGQKE-QQOCEESCKS--QYGEKDDQ 45
 DB 147 QDCQACPPQOQPOQOQOQOQOQOQOQOQOQOQOQ 185

RESULT 9

S50830
 Machado-Joseph disease MJD1a protein - human

C:Species: Homo sapiens (man)
 C:Date: 14-Jul-1995 #sequence_revision 21-Jul-1995 #text_change 28-May-1999
 C:Accession: S50830
 R:Kawaguchi, Y.; Okamoto, T.; Taniwaki, M.; Aizawa, M.; Inoue, M.; Katayama, S.; Kawakami
 Nature Genet. 8, 221-228, 1994
 A:Title: CAG expansions in a novel gene for Machado-Joseph disease at chromosome 14q32.1
 A:Reference number: S50830; MUID:95179166
 A:Accession: S50830
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-360 <RAW>
 A:Cross-references: GB:S75313; NID:9833927; PIDN:AAB33571.1; PUID:9833928

Query Match 22.5%; Score 62; DB 2; Length 360;
 Best Local Similarity 35.0%; Pred. No. 11;
 Matches 14; Conservative 12; Mismatches 12; Indels 2; Gaps 1;

QY 10 EDCRRRCCEMDTRGQKEQOQCEESCKSQYGEKDDQ 47
 DB 279 EELRRKEAFAFEKQKQOQOQOQOQOQOQOQOQOQ 318

RESULT 10
 F64533

co-chaperone and heat shock protein - Helicobacter pylori (strain 26695)

C:Species: Helicobacter pylori
 C:Date: 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 08-Oct-1999
 C:Accession: F64533
 R:Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.
 Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKee
 son, J.D.; Kelley, J.M.; Cotton, M.D.; Feldman, J.M.; Fujii, C.; Bowman, C.; Watthey,
 Nature 388, 539-547, 1997
 A:Authors: Mallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser,
 A:Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.
 A:Reference number: A64520; MUID:97394467
 A:Accession: F64533
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-189 <TOM>

A:Cross-references: GB:AE000532; GB:AE000511; NID:g231384; PIDN:AAD07179.1; PID:g231
 C:Superfamily: heat shock protein gripe

Query Match 22.2%; Score 61; DB 2; Length 189;
 Best Local Similarity 44.4%; Pred. No. 7.8;
 Matches 12; Conservative 5; Mismatches 10; Indels 0; Gaps 0;

QY 17 EMDTRGQKEQOQCEESCKSQYGEKDDQ 43
 DB 8 EHDHLSQKEPECEKCKEQYERKE 34

RESULT 11
 A33513

hypothetical protein B - human T-cell lymphotropic virus type 1

C:Species: human T-cell lymphotropic virus type 1, HTLV-1
 C:Date: 20-Dec-1989 #sequence_revision 20-Dec-1989 #text_change 18-Jun-1993
 C:Accession: A33513
 R:LaRocca, D.; Chao, L.A.; Seto, M.H.; Brunck, T.K.
 Biochem. Biophys. Res. Commun. 163, 1006-1013, 1989
 A:Title: Human T-cell leukemia virus minus strand transcription in infected T-cells.
 A:Reference number: A33513; MUID:89391952
 A:Accession: A33513
 A:Status: preliminary; not compared with conceptual translation
 A:Molecule type: genomic RNA
 A:Residues: 1-265 <LAR>

Query Match 22.2%; Score 61; DB 2; Length 265;
 Best Local Similarity 26.5%; Pred. No. 11;
 Matches 13; Conservative 17; Mismatches 17; Indels 2; Gaps 1;

QY 1 GDDDPK--RYEDCRRRCCEMDTRGQKEQOQCEESCKSQYGEKDDQ 47
 DB 123 GEKAPRGETHRDQRRAEKKKKKKKKKKKKKKKKKKKK 171

RESULT 12
 T18267

multidrug resistance protein - slime mold (Dictyostelium discoideum)

C:Species: Dictyostelium discoideum
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
 C:Accession: T18267
 R:Shaulsky, G.; Kuspa, A.; Loomis, W.F.
 submitted to the EMBL Data Library, January 1995
 A:Description: An MDR transporter/serine protease gene is required for prestalk speci
 A:Reference number: Z18850
 A:Accession: T18267
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-1905 <SHA>

A:Cross-references: EMBL:U20432; NID:9664839; PID:g664840; PIDN:AAA62212.1
 C:Genetics:
 A:Gene: tagB

Query Match 22.2%; Score 61; DB 2; Length 1905;
Best Local Similarity 28.3%; Pred. No. 60;
Matches 13; Conservative 13; Mismatches 20; Indels 0; Gaps 0;

OY 2 DDDPKRYEDCRRRCCEWDTGQKEQOQCEESCKSQYGEKDQOQRHR 47
DB 1805 DDDPKRYEDCRRRCCEWDTGQKEQOQCEESCKSQYGEKDQOQRHR 1850

RESULT 13

T21688

hypothetical protein F33A8.1 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C:Accession: T21688

R:Matthews, L.

submitted to the EMBL Data Library, November 1996

A:Reference number: Z19459

A:Accession: T21688

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-897 <M11>

A:Cross-references: EMBL:Z81525; PIDN:CAB04256.1; GSPDB:GN00020; CESP:F33A8.1

A:Experimental source: clone F33A8

A:Gene: CESP:F33A8.1

A:Map position: 2

A:Introns: 29/3; 56/3; 98/2; 225/2; 384/3; 419/1; 488/2; 517/3; 637/2

Query Match 22.0%; Score 60.5; DB 2; Length 897;
Best Local Similarity 31.1%; Pred. No. 35;
Matches 14; Conservative 15; Mismatches 15; Indels 1; Gaps 1;

OY 3 DDDPKRYEDCRRRCCEWDTGQKEQOQCEESCKSQYGEKDQOQRHR 47
DB 797 DDDPKRYEDCRRRCCEWDTGQKEQOQCEESCKSQYGEKDQOQRHR 840

RESULT 14

A71928

cag island protein - Helicobacter pylori (strain J99)

C:Species: Helicobacter pylori

A:Variety: strain J99

C:Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 08-Oct-1999

C:Accession: A71928

R:Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R.;

; Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.;

Nature 397, 176-180, 1999

A:Title: Genomic sequence comparison of two unrelated isolates of the human gastric path

A:Reference number: A71800; MUID:99120557

A:Accession: A71928

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1819 <AARN>

A:Cross-references: GB:AE001481; GB:AE001439; NID:94155005; PIDN:AA06047.1; PID:9415500

A:Experimental source: strain J99

C:Genetics:

A:Gene: orf13/14

Query Match 22.0%; Score 60.5; DB 2; Length 1819;
Best Local Similarity 34.1%; Pred. No. 65;
Matches 15; Conservative 9; Mismatches 13; Indels 7; Gaps 2;

OY 7 KRYEDCRRRCCEWDTGQKEQOQCEESCKSQYGEKDQOQRHR 47
DB 1242 KRYEDCRRRCCEWDTGQKEQOQCEESCKSQYGEKDQOQRHR 1281

RESULT 15
H81536

hypothetical protein CP0796 [imported] - Chlamydia pneumoniae (strain AR39)

C:Species: Chlamydia pneumoniae

C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 18-Aug-2000

C:Accession: H81536

R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hicke,

, C.; Dodson, R.; Gwin, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzbe

r, Nucleic Acids Res. 28, 1397-1406, 2000

A:Title: Genome sequences of Chlamydia trachomatis M0p1 and Chlamydia pneumoniae AR39

A:Reference number: A81500; MUID:20150255

A:Accession: H81536

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-241 <REA>

A:Cross-references: GB:AE002239; GB:AE002161; NID:97189708; PIDN:AAF38595.1; PID:9718

A:Experimental source: strain AR39, HL cells

C:Genetics:

A:Gene: CP0796

C:Superfamily: Chlamydia pneumoniae hypothetical protein CP0796

Query Match 21.8%; Score 60; DB 2; Length 241;
Best Local Similarity 26.2%; Pred. No. 12;
Matches 11; Conservative 12; Mismatches 19; Indels 0; Gaps 0;

OY 2 DDDPKRYEDCRRRCCEWDTGQKEQOQCEESCKSQYGEKDQO 43
DB 51 DDDPKRYEDCRRRCCEWDTGQKEQOQCEESCKSQYGEKDQO 92

Search completed: March 1, 2001, 15:52:44
Job time: 568 sec